DOCUMENT RESUME

ED 087 159 EC 061 189

AUTHOR Moonan, William J.
TITLE Charosel: A Computer Program Which Selects Qualitative Predictors for Qualitative Criterion Prediction Problems.
PUB DATE 73
NOTE 30p.; Paper presented at Military Testing Association Conference (San Antonio, Texas, October 28 through November 2, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Computer Programs; *Exceptional Child Research; Factor Analysis; *Gifted; *Identification; Kindergarten Children; *Statistical Analysis; Theories

ABSTRACT

Described is the theory and computer program of a new statistical procedure to find a parsimonious subset of predictor variables which have high predictive efficiency for a given criterion variable, and explained is application of the procedure to problems such as the identification of gifted kindergarten children. The procedure is seen to be most appropriate for Type D prediction problems, for which the criterion variable is represented by a set of categories and the predictor variables are represented by a set of qualitative variables. It is explained that the technique is based on a formula for calculating a posteriori probabilities and a simple decision function, and that models are evaluated by computing decision tables and an associative objective function for each predictor variable available. The program is thought to be applicable to prediction problems of the military, education, and medicine. Described is application of the procedure to the selection of questionnaire items most predictive of giftedness in kindergarten children. (DB)

BEST COPY AVAILABLE

ED 087159

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

CHAROSEL: A COMPUTER PROGRAM WHICH SELECTS QUALITATIVE PREDICTORS FOR QUALITATIVE CRITERION PREDICTION PROBLEMS

by

William J. Moonan, PhD

Director of Statistical Programs
Navy Personnel Research and Development Center (NPRDC)
San Diego, California 92152

A Presentation Made to

The 1973 Military Testing Association Conference
San Antonio, Texas

28 October - 2 November 1973

EC 061189

SUMMARY

A. Problem

To develop the theory and computer program to effect model searching procedures for Type D prediction problems.

B. Background and Requirements

A Type D prediction problem is one for which the criterion variable is represented by a set of categories and the predictor variables are represented by a set of qualitative variables. Model searching is associated with the process of finding a parsimonious subset of predictor variables for which little predictive efficiency is lost by using this subset as compared to the efficiency of prediction found by using the entire available predictor set.

C. Approach

The technique used to solve this problem is based on Bayes' formula for calculating a posteriori probabilities and a simple decision function. Models are evaluated by computing decision tables and an associative objective function for each predictor variable available from a testing sample of observations. The fundamental theory is based on the so-called ABCD-Technique the author presented at the 1972 Military Testing Association Conference.

D. Findings

The program and procedure is applicable to many military personnel research problems as well as to the prediction problems of education and medicine. The program has been applied to several important

personnel research problems with excellent success, frequently finding models with percent correct classification accuracy greater than 70% when utilizing less than 20% of the available predictor information.

E. Conclusion

The theory and the computer program for the CHAROSEL technique have been validated and applied successfully to NPRDC and other personnel research projects.

F. Recommendations

It is recommended that the CHAROSEL technique be applied to and evaluated for other naval personnel research projects and be made available upon request to other Department of Defense agencies.

TABLE OF CONTENTS

	Page
Summary	iii
I. INTRODUCTION	1
A. Purpose	1
B. Background	1
1. What is the kind of problem?	2
a. The discrimination problem	3
b. The taxonomy problem	3
c. The classification or diagnosis problem	3
d. Types of variables	4
2. Problem statement	5
II. THEORY	5
A. Notation	5
B. Procedure Description	6
1. General	6
2. Variable selection technique	7
3. CHANCELL's objective function	7
III. IMPLEMENTATION	9
A. The CHAROSEL Technique: Computer Program Documentation	9
B. Identification	9
C. Usage	10
1. Input deck setup	10
2. Input card layout	10
3. Equipment configuration	10
4. Output	10
D. Requirements	11
1. Storage	11
2. Time	11
E. Program Restrictions	11
F. User Options	12

Page

G. An Illustration 13

1. The problem description 13
2. The results 14

IV. CONCLUSION 16

Acknowledgments 16

References 18

Addendum 17

LIST OF TABLES

Table		Page
1.	Types of Prediction Analysis	20
2.	Illustrative Decision Table for a Three Category Problem . .	21
3.	CHAROSEL Decision Table for First Stage Analysis for the Gifted Children Problem	22
4.	CHAROSEL Decision Table for Fourth Stage Analysis for the Gifted Children Problem	23
5.	Recent NPRDC CHAROSEL Applications	24

I. INTRODUCTION

A. Purpose

The author realized that statistical theory and practice had a serious gap; namely, that no adequate statistical technique existed for the analysis problem associated with the prediction of categorical membership from knowledge of other categorical (qualitative) information. He attempted (Moonan, 1972) to provide an adequate solution to this problem. His solution, augmented by a computer program, permitted the utilization of possibly several hundred available qualitative predictors. Since most operational situations employing prediction analysis cannot practically utilize predictor information of this magnitude, it was soon realized that some model seeking technique would have to be developed in order to make the ABCD technique more effective. Other prediction analysis procedures have their associative model seeking procedures as illustrated in Table 4, p. 17, of Moonan (1973). These other variable selection techniques were not useful for problems analyzed by the ABCD technique and, consequently, a new procedure called CHOOSEL was developed.

B. Background

The initiation of this development began with a real personnel problem, namely that of classifying individuals into various categories of drug abusers from knowledge of biographical, life-history, and opinion data items. It may be of interest to this

audience to note that both the ABCD and CHAROSEL procedures were conceived, analyzed, and programmed before data from the project were made available for analysis. There are several lessons to be learned from this experience, none of which can be elaborated here, other than the bemusing consideration that, contrary to many research analysts' opinion, personnel research planning and data collection activities are more time consuming than data analysis: Personnel research planning and data collection activities may indeed be much more time consuming than the development of data analysis techniques.

The type of prediction problem referred to has been labeled by the author as a Type D prediction problem. Other models for prediction analysis have been identified by Lubin (1950) and Rozeboom (1966) and are labeled as in Table 1.

Insert Table 1 about here

---

1. What is the kind of problem? Statistical consultants and analysts frequently find it useful to answer the question, "What kind of problem is this?" when consulting with clients or developing analysis procedures. The procedure has several advantages, since, if the question can be answered succinctly and clearly, the complications associated with the consultation or analysis are greatly simplified. Accordingly, we shall attempt to use this technique in our description. The answer to the question is most

appropriately the identification of a particular class of statistical classification techniques. Some of the candidates are indicated as follows:

a. The discrimination problem. Given a set of distinct populations and a multivariate random variable, the discrimination problem is that of defining a numerical index using the multivariate information in such a way that this index optimally discriminates among the populations. The classical solution to this problem is given by the linear discriminant function of R. A. Fisher (1936). The appropriate model seeking technique for this problem utilizes Mahalanobis' generalized distance statistic (1936) as illustrated by Rao (1952).

b. The taxonomy problem. If the researcher has reason to believe that his multivariate sample came from more than one distinct population, the taxonomy problem is used to evolve a rule for sorting the sample observations into several distinct populations. An approach to this problem solution for multivariate normal predictors has been provided by Wolfe (1967) and for simple binary attribute predictors by Tanimoto (1960). No specific model seeking procedures have been developed for these techniques or are known to the author.

c. The classification or diagnosis problem. Given a group of distinct multivariate populations for which samples are available, the solution to the classification problem is a decision rule that would enable one to assign an additional observation to his proper

population if the multivariate observational vector is available and it is known that the individual has been derived from one of the given populations. There are a variety of techniques available for application, all of which are based, in one way or another, on modern decision theory and, for the most part, utilize multivariate normal assumptions. Among these solutions are the Wald (1944) statistic and the Anderson (1951) approach. Model seeking approaches within the classification problem solutions are not generally available.

d. Types of variables. The concept of variables enter into the problem statement in two different ways. The first way occurs when consideration is given to the fact that we have both a criterion variable, i.e., category to be properly identified for an individual and a set of predictor variables from which the criterion category is to be ascertained. It is also interesting to note that our problem is uniquely concerned with variables that are described as "qualitative" rather than "quantitative." These types of variables differ with respect to the way by which their measurement is "manifested." Qualitative variable measurements are manifested by a specification of a quality of the variable on a nominal scale of measurement. When the variable is manifested by or for an object through an item of a questionnaire, the response quality is called an alternative. Quantitative variable measurement is manifested on an interval or ratio scalable, metrical dimension as a quantity. Our concern, in this report, is exclusively with qualitative variables used both as criteria and predictors.

From these considerations it definitely appears that the problem at hand is a classification problem whose solution, because of the qualitative nature of the predictor information, is provided by Moonan's ABCD technique for which, until now, no model seeking procedure was available.

2. Problem statement. Thus our problem, succinctly stated, is to devise a model selection technique for prediction problems involving the prediction (or decision) of the category of membership using as an information base for prediction a large set of other qualitative variables.

II. THEORY

A. Notation

θ_k represents the k^{th} category of group membership, $k = 1(1)K$.

(1) $X_i = (x_{1i}, x_{2i} \ldots, x_{mi})$ represents the predictor qualitative variables associated with individual i.

(2) $P(X_i, \theta_k)$ represents the probability that individual i possesses predictor information X_i and is a member of criterion category k.

(3) $P(\theta_k)$ represents the a priori probability of a randomly selected individual being a member of category k.

(4) $P(x_m^a \mid \theta_k)$ represents a probability, known as an endorsement ratio, indicating the probability of an individual from category k responding to the a^{th} alternative of the m^{th} predictor character.

(5) Then $P(X_i \mid \theta_k) = \frac{P(X_i, \theta_k)}{P(\theta_k)}$ represents the conditional probability of X_i given θ_k. We are, however, primarily interested in the "inverse" of this probability which is obtained from Bayes' (1763) formula expressed as follows:

$$(6)\quad P(\theta_k \mid X_i) = \frac{P(\theta_k)P(X_i \mid \theta_k)}{\sum_{k=1}^{K} P(X_i)P(X_i \mid \theta_k)} .$$

With this notation we are now in a position to describe the theoretical process of our Bedel selecting procedure called CHAROSEL.

B. Procedure Description

1. General. The CHAROSEL procedure computes a vector of the Bayes' formula probabilities, $P(\theta_k \mid X_i)$, for each member of a testing sample for which θ_k is known. This vector is vector multiplied by the column elements of a cost of misclassification matrix $C_{\theta,\theta'}$. The elements of $C_{\theta,\theta'}$ equal zero when $\theta = \theta'$. Otherwise, $C_{\theta,\theta'} > 0$. This vector multiplication produces a set of K expected misclassification costs, V_k, and the decision is that the i^{th} individual is a member of category θ_e if V_e is the minimum of the V_k.

In order to maintain computational feasibility, it is assumed that the elements of X_i are stochastically independent so that:

$$(7)\quad P(X_i, \theta_k) = \prod_{m=1}^{M} P(X_{im}, \theta_k) .$$

2. Variable selection technique. We shall accomplish the model seeking procedure by characterizing it as a problem of selection of the qualitative predictor variables. A similar approach to this class of problems has been frequently used by the author (Moonan, 1956, 1967, 1968). The optimal approach to this form of solution has long been considered by the author to be that of dynamic programming similar to the extremely clever solution to the clustering problem provided by Jensen (1968). Unfortunately, the dynamic programming solution to the model seeking problem has eluded me and I have had to be content, as in this case, with using what I call "conditional sequential optimization." This optimization procedure is defined as follows: The program selects predictors in such a manner as to optimize a meaningful objective function. The variable is first selected which, by itself, has the best objective function value; next, another variable is selected such that, when combined with the first selected variable, the two produce a predictor system with best conditional objective function value. Having found that pair, a new third predictor is sought which, when combined with the previous selected pair, produces a three variable predictor system with the best conditional objective function value. This process is repeated until all predictor variables are utilized.

3. CHAROSEL's objective function. In order to discuss the objective function of CHAROSEL, it is necessary to consider Table 2, which portrays the results from applying the ABCD technique for any

Insert Table 2 about here

given number of predictors in a three criterion category prediction problem. Table 2 may be interpreted as follows: There are three criterion categories labeled 1, 2, and 3, for which testing sample sizes of 39, 24, and 72, respectively, are available. Using AEOD for some specified set of predictor variables, 30 of category 1, 17 of category 2, and 61 of category 3 were correctly classified.

If the predictor variables were completely effective, the principle main - diagonal (30, 17, 61) would be represented by (39, 24, 72), showing that all testing sample members were correctly classified and all non-diagonal elements would have frequency = 0.

This suggests that the proper objective function is a statistic representing this condition. One such statistic is provided by an application of classical information theory and was suggested by the author to his programmer, Mr. Samuel Bowser. This statistic proved inappropriate for some applications and was modified by Mr. Bowser to equation (8).

(8) Objective Function

$$\frac{-\sum_{i=1}^{k}\left\{Pi - \sum_{\substack{j=1 \\ i\neq j}}^{k} P_{ij}\,\mathrm{Log}_2(P_{ij})\right\} + \sum_{i=1}^{k}\frac{n_{ii}}{f_{ii}}}{k} \quad \mathrm{Lim} \to 1.$$

$$\lim_{f_{ii} \to 0} \frac{n_{ii}}{f_{ii}} = n_{ii}$$

P_i = A priori probability.

P_{ij} = Cell proportion off diagonal

n_{ii} = Number of occurrences in a given group

f_{ii} = Diagonal cell frequency

k = Number of groups

$\lim_{x \to 0} x \log x = 0$

Equation (8) has a value = 1.00000 if the decision tables' diagonal equals the frequencies of the testing samples' category sizes, otherwise the statistic is greater than unity.

III. IMPLEMENTATION

A. The CHAROSEL Technique: Computer Program Documentation

The CHAROSEL technique was programmed as a computer program for the NELC IBM 360/65 system and can be executed through the NPRDC IBM 1130 remote job entry system. The documentation for this program follows as extracted from Bowser (1973) and Noonan and Bowser (1973).

B. Identification

Title: CHAROSEL

Common Identification: CHARI

Category: Statistical

Programmer: Samuel E. Bowser

Analyst: William J. Moonan

Date: 1 May 1973

Compiler: IBM 360/65 FORTRAN G

Core Requirements: 375K

C. Usage

1. Input deck setup

a. Program deck and subroutine decks.

b. Input form including parameter and option cards.

c. Data either card or 9-track tape.

2. Input card layout

See input forms, Bowser (1973).

3. Equipment configuration

Input form cards are always loaded behind the program deck on the standard input unit. Input data can be on a separate tape or loaded from the standard input unit.

4. Output

a. Printer

(1) Decision table as each variable is selected.

(2) Values of objective function, change in objective function, total per cent correct decisions, and change in percentages of correct decisions.

(3) Variables are listed in rank order as they are selected.

b. Punch

(1) Endorsement ratios are punched out if they are calculated by program; this option is selected unless otherwise specified.

D. Requirements

1. Storage. The program as currently constructed does not utilize object time dimensioning. The size of the core requirement at present is 375K. This allows for 500 men and 200 variables. The number of categories is fixed at 10 or less and the levels of the variables is fixed at 10 or less. In order to exceed the size restrictions the redimensioning of the main program and its subroutines would be necessary.

2. Time. An approximation of the time required can be obtained from the following:

$$\text{Time (Min.)} = 5 + \left| \frac{(\text{Nr. of Variables})^2}{300} - 14 \right| (\text{Nr. of Categories} - 1)$$

Thus, for a three category, 50 predictor variable problem, the computing time of CHAIDSEL for normal options is about 17 minutes.

For many other important details regarding documentation and instructions for filling out the input forms the user should consult Bowser (1972).

E. Program Restrictions

1. The training sample size must be less than or equal to 500.

2. The number of criterion categories must be less than or equal to 10.

3. The number of levels for each predictor character must be less than or equal to 10.

4. The number of predictors is limited by computer system core size and is related to sample size. Adjustment may be made to larger numbers of predictors but at present the program is set for 200 predictor variables maximum.

5. At present the program is operational only on an IBM 360/65 computer system and utilizes a core size of 300K bytes.

6. Prospective DOD users may arrange for program execution through:

Head, Computer Services Department

Navy Personnel Research and Development Center

San Diego, California 92152

7. Other potential users should contact the authors.

F. User Options

The user of CHAROSEL has considerable freedom with regard to type of procedure he wishes to utilize with this program. Among these options are:

1. The endorsement ratios may be either supplied by the user or the program will calculate them from the training sample.

2. A limit may be given for the maximum number of variables with missing data that are allowed for each object of the training sample

and/or missing data may be used as a "level" in the corresponding predictor variable.

3. Certain predictor variables may be forced to be utilized initially in the CHAROSEL predictor variable identification at the option of the user.

4. If the user wishes to utilize the cross validation process this may be accomplished by taking the variables selected by CHAROSEL and running them on the AFCD program with a testing sample.

5. Training sample data may be either on cards or on tape, however, there exists an input form to be completed by the user before execution.

G. An Illustration

1. The problem description. As an illustration of the CHAROSEL program, we present a problem actually executed on the program and provide the output results obtained thereby. This illustration is concerned with the educational problem of identifying intellectually gifted children who are members of so-called low socio-economic status groups within the city of San Diego. This illustration is the PhD thesis problem of Mrs. Charlotte Malone at the United States International University, San Diego. As yet, the thesis is incomplete and, of course, unpublished, and the reference to this work is only provided for illustrative purposes. The categories of criterion for this problem are two: θ_1, the children who have Stanford Binet IQ's greater than or equal to 132 and labeled "gifted" and, θ_2, those

children who have Stanford Binet IQ's less than 132 who are labeled "non-gifted."

The predictor information was derived from a questionnaire completed by the mothers of kindergarten children registering for school within a certain school district of San Diego during September 1973. The items of the questionnaire were composed of 51 biographical, behavioral, and life history items such as:

(1) Does your child use a lot of "grown-up" words? Does he talk like an older person than he really is?

A. Seldom

B. Occasionally

C. Often

D. Almost always

E. No answer possible

2. The results. The training sample for this problem is at present very small but will be augmented within a short time. For the present, $N_1 = 28$ gifted students and $N_2 = 21$ non-gifted students on whom data were available. The researcher specified the a priori probabilities to be $p(\theta_1) = .05$ and $p(\theta_2) = .95$. Each item of the questionnaire had five levels of response and their endorsement ratios, for each category, were computed by CHANOSEL. The execution of the program produced the following results: (See Table 3).

Insert Table 3 about here

We note that the decision table showed 17 misclassifications of gifted children and no misclassifications of non-gifted children and that the most optimal single predictor item was item number 1, shown previously. The objective function value was 1.78957 and per cent correct accuracy = (32/49)100 = 65.31.

The program proceeded to sequentially select items 29, 7, and 27, producing the information outlined in Table 4.

Insert Table 4 about here

At this stage, only eight misclassifications were made and the objective function has been reduced to 1.21067 and the per cent correct accuracy = $(\frac{41}{49})100 = 83.67$.

After 30 stages of selection the main - diagonal of the decision table was (28, 21), indicating that the per cent correct accuracy of classification was 100.00. The statistical results for this stage of analysis are available but not shown in this report.

The author has experienced the greatest personal satisfaction as a result of the utilization of his CHAROSEL technique to this identification problem which, if these preliminary results are sustained, will lead to better educational practices and improved well-being for many children.

The CHAROSEL program has been applied to several military and civilian personnel research problems. Some of these results are shown in Table 5.

Insert Table 5 about here

IV. CONCLUSION

The CHAROSEL technique represents a unique development in classification theory in that it possibly provides the only practical procedure available for selecting effective qualitative predictor variables for Type D prediction problems. A computer program, associated with this technique, has been developed and applied to several problems associated with both military and civilian personnel research. The results from these applications have shown the technique can identify a few predictor variables which will produce results of very high classification efficiency. This report provided information relative to the theory, implementation, and application of the program.

ACKNOWLEDGMENTS

The author would like to thank:

(1) Mr. Edmund Thomas, whose research on an NPRDC drug abuse project suggested the requirement for developing the AECD technique.

(2) Dr. Norman Abrahams, whose work with program KEYCON and personal discussions suggested the need for the CHAROSEL program.

(3) Mr. Samuel E. Bowser for efficiently programming the CHAROSEL technique on the 360/65 computer.

(4) Mrs. Teddy Ralph, Mrs. Delores Callahan, and William K. Moonan for their assistance in writing this report.

(5) Mrs. Charlotte Malone for allowing me to discuss her PhD thesis problem and to present some data relating to it.

ADDENDUM

After this report was completed, the following thoughts occurred to the author.

The controlling aspect of CHAROSEL is its objective function. Suppose this was modified from equation (8) to the bivariate uncertainty measure

$$U(.,.) = - \sum_{i=1}^{k} \sum_{j=1}^{k} p(i,j) \log p(i,j) .$$

where $p(i,j)$ is the joint probability of row i and column j in the CHAROSEL decision table. The objective of the procedure is, then, to select variables that maximize joint uncertainty, $U(.,.)$. Used in this way the program could be used to select variables to affect heterogeneity among categories in the sense that the program will identify those variables which tend to make an independent relationship between predicted category and actual category.

This approach appears to have application to some analysis problems encountered at NPRDC.

REFERENCES

Anderson, T. W. (1951). Classification by multivariate analysis. Psychometrika, 16, 31-50.

Bayes, T. R. (1763). An essay towards solving a problem in the doctrine of chances. The Philosophical Transaction, Vol. LIII, 370-418.

Bowser, S. E. (1973). CHAROSEL I. Unpublished. Naval Personnel and Training Research Laboratory Statistical Analysis and Computer Applications Department.

Fisher, R. A. (1936). Use of multiple measurements on taxonomic problems. Annals of Engineers, Vol. VII, 179-188.

Jensen, R. E. (1968). A dynamic programming algorithm for cluster analysis. Management Science, 14, 1034-1057.

Lubin, A. (1950). Linear and non-linear discriminating functions. British Journal of Psychology, Section 3, 90-103.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. Proceedings of the National Institute of Science (India), Vol. 12, 49-55.

Moonan, W. J. (1972). ABCD: A Bayesian technique for making discriminations with qualitative variables. Proceedings of the 1972 Military Testing Association Conference.

Moonan, W. J. (1973). Vistas of analysis: Qualitative prediction and decision analysis. Proceedings of the 1973 Occupational Research Conference at the Navy Personnel Research and Development Center, San Diego.

Moonan, W. J., Balaban, J. G., Geyser, M. J. (1967). SEQUIN II: A computerized item selection and regression analysis procedure. U. S. Naval Personnel Research Activity. (Research Memorandum SRM 68-11)

Moonan, W. J., Beauregard, C. (1968). PARFORM: A computer program for developing statistically parallel tests of high validity. San Diego: U. S. Naval Personnel Research Activity. (Research Memorandum SRM 68-22)

Moonan, W. J., & Bowser, S. E. (1973). CHAROSEL Synopsis. Fact Document #27, Naval Personnel & Training Research Laboratory, San Diego.

Moonan, W. J., & Pooch, V. W. (1966). SEQUIN: A computerized item selection procedure. San Diego: U. S. Naval Personnel Research Activity. (Research Memorandum SRM 67-8)

Rao, C. R. (1952). Advanced statistical methods in biometric research. New York: John Wiley and Sons.

Rozeboom, J. H. (1966). The foundations of the theory of prediction. Homewood, Ill.: The Dorsey Press.

Tanimoto, T. T., & Rodgers, D. J. A computer program for classifying plants. Science, 1960, 132, 1115-1123.

Wald, A. (1944). On the statistical problem arising in the classification of an individual into one of two groups. Annals of Mathematical Statistics, 15, 145-162.

Wolfe, J. H. (1967). NORMIX: Computer program documentation. San Diego: U. S. Naval Personnel Research Activity. (Research Memorandum SRM 69-11)

TABLE 1

Types of Prediction Analysis

Independent Variables	Dependent Variable	
	Metrical Criteria	Categorical Criteria
Metrical Predictors	Regression Analysis Type A	Discriminant Analysis Type C
Categorical Predictors	Analysis of Variance Type B	ABCD Analysis Type D

TABLE 2

Illustrative Decision Table for a Three Category Problem

True Categories	Decision Category 1	2	3	Total	Percent
Category 1 Frequencies	30	4	5	39	28.9
Category 2 Frequencies	4	17	3	24	17.8
Category 3 Frequencies	6	5	61	72	53.3
Column Totals	40	26	69	135	
Column Percents	29.6	19.3	51.1		100

TABLE 3

CHAROSEL Decision Table for First Stage Analysis
For the Gifted Children Problem

True Categories	Decision Category 1	2	Total	Percent
Category 1 Frequencies	11	17	28	57
Category 2 Frequencies	0	21	21	42
Column Totals	11	38	49	
Column Percents	22	77		

Stage Selected	Variable Name	Nr.	Objective Value	Function Change	Percent Amount	Correct Change
1	01	1	1.78597	0.0	65.31	0.0

TABLE 3

CHAROSEL Decision Table for First Stage Analysis
For the Gifted Children Problem

True Categories	Decision Category 1	2	Total	Percent
Category 1 Frequencies	11	17	28	57
Category 2 Frequencies	0	21	21	42
Column Totals	11	38	49	
Column Percents	22	77		

Stage Selected	Variable Name	Nr.	Objective Value	Function Change	Percent Amount	Correct Change
1	01	1	1.78597	0.0	65.31	0.0

TABLE 4

CHAROSEL Decision Table for Fourth Stage Analysis for
The Gifted Children Problem

True Categories	Decision Category 1	2	Total	Percent
Category 1 Frequencies	20	8	28	57
Category 2 Frequencies	0	21	21	42
Column Totals	20	29	49	
Column Percents	40	59		

Stage Selected	Variable Name	Nr.	Objective Value	Function Change	Percent Amount	Correct Change
1	01	1	1.78597	0.0	65.31	0.0
2	29	29	1.38743	0.39855	75.51	10.20
3	7	7	1.24807	0.13936	81.63	6.12
4	27	27	1.21067	0.03740	83.67	2.04

TABLE 5

Recent NPRDC CHAROSEL Applications

Problem Name	Criterion Categories	Types of Predictors	Training Sample	Nr. of Predictors Used	Objective Function Value	% Classification Accuracy
Drug Abuse Prediction	(3) A. Hard Drug User B. Recreational Drug User C. Abstainers	(133) A. Biographical B. Opinion Data	(362) Naval Recruits	(5) (18)	1.[illegible]3207 1.72074	81.6 68.2
Odds for Effectiveness	(2) A. Recommended for Reenlistment B. Not Recommended	(140) A. Biographical B. Self-evaluation C. Opinion	(330) Naval Recruits	(7) (18) (95)	1.38470 1.11748 1.00000	95.6 98.4 100.0
Skid Row Deployment	(2) A. Completed Project B. Dropped Out	(52) A. Biographical B. Self-evaluation	(172) Skid Row Welfare Recipients	(13) (39)	1.38013 1.26212	85.5 88.8
Target Classification	(3) A. U. S. Subs B. Foreign Subs C. Surface Ships	(10) Sonar Class	(91) Sonar Contacts	(1) (4) (10)	1.51500 1.07334 1.07334	70.3 95.6 95.6
Identification of Gifted Children	(2) A. Gifted Children B. Non-Gifted Children	(51) Behavioral Characteristics Questionnaire	(49) School Children in Kindergarten	(1) (4) (30)	1.78957 1.31067 1.00000	65.31 83.67 100.00